Chaos

Counting forbidden patterns in irregularly sampled time series. I. The effects of under-sampling, random depletion, and timing jitter

Cite as: Chaos **26**, 123103 (2016); https://doi.org/10.1063/1.4968551
Submitted: 08 September 2016 . Accepted: 21 October 2016 . Published Online: 06 December 2016

Michael McCullough, Konstantinos Sakellariou, Thomas Stemler, and Michael Small

CHAOS 26, 123103 (2016)

Counting forbidden patterns in irregularly sampled time series. I. The effects of under-sampling, random depletion, and timing jitter

Michael McCullough,[1,a)] Konstantinos Sakellariou,[1] Thomas Stemler,[1,2] and Michael Small[1,2,3]
[1]*School of Mathematics and Statistics, The University of Western Australia, Crawley, Western Australia 6009, Australia*
[2]*Complex Data Modelling Group, Faculty of Engineering, Computing and Mathematics, The University of Western Australia, Crawley, Western Australia 6009, Australia*
[3]*Mineral Resources, CSIRO, Kensington, Western Australia 6151, Australia*

(Received 8 September 2016; accepted 21 October 2016; published online 5 December 2016)

It has been established that the count of ordinal patterns, which do not occur in a time series, called forbidden patterns, is an effective measure for the detection of determinism in noisy data. A very recent study has shown that this measure is also partially robust against the effects of irregular sampling. In this paper, we extend said research with an emphasis on exploring the parameter space for the method's sole parameter—the length of the ordinal patterns—and find that the measure is more robust to under-sampling and irregular sampling than previously reported. Using numerically generated data from the Lorenz system and the hyper-chaotic Rössler system, we investigate the reliability of the relative proportion of ordinal patterns in periodic and chaotic time series for various degrees of under-sampling, random depletion of data, and timing jitter. Discussion and interpretation of results focus on determining the limitations of the measure with respect to optimal parameter selection, the quantity of data available, the sampling period, and the Lyapunov and de-correlation times of the system. *Published by AIP Publishing.*
[http://dx.doi.org/10.1063/1.4968551]

Experienced practitioners of nonlinear time series analysis will understand the importance discerning deterministic dynamics from stochastic processes using measured data. The reasons are, first, that many of the existing techniques in the field are theoretically dependent on the assumption of determinism. This includes any method requiring a phase space reconstruction using time delay embedding such as the estimation of the correlation dimension or Lyapunov exponents.[1] Second, knowing from experimental data that a system that is likely driven by a deterministic process can be used to justify and inform the construction of models as systems of equations. The count of forbidden ordinal patterns[2] is a relatively new measure that has been shown to be a powerful and computationally simple tool for the detection of determinism in severely noise corrupted data, but how reliable is this measure when the time series data have been sampled with insufficient frequency or at irregular intervals?

I. INTRODUCTION

Irregularly sampled time series are commonplace in data that have been measured beyond the confines of controlled laboratory experiments. More extreme examples can arise where measurements can only be taken in particular weather conditions or when some remote sensing apparatus is within the range of communication such as astrophysical data,[3] which must be recorded with telescopes or received from satellites. There are also data that are inherently irregularly sampled such as financial time series where stock prices are updated when transactions are made,[4] or recordings of biological signals such as cardiac inter-beat intervals.[5,6] A further example is time series affected by the timing jitter in measurement devices, which can induce bias in amplitude measurements of sampled waveforms.[7]

The reality of irregular sampling compels us to investigate the reliability of any measure or method that we would apply to analyse such data. In a very recent paper by Kulp *et al.*,[8] the authors undertook a study of the robustness of the count of forbidden patterns with respect to several irregular sampling schemes. Forbidden patterns are a phenomenon that arises from the process of mapping deterministic time series to a particular symbolic dynamics known as ordinal patterns.[2,9] This method of abstraction was originally introduced by Bandt and Pompe to quantify complexity,[10] but subsequent research has shown that identifying the existence of a set of ordinal patterns that do not occur in a given time series, or in the language of the field, confirming the existence of some set of forbidden patterns, is a theoretically sound and practically applicable criterion for asserting that the data contain a deterministic component.[2,11,12] Kulp *et al.* used numerically generated time series from periodic and chaotic regimes of the Lorenz system to test the effects of irregular sampling on the count of forbidden patterns and found that the measure was robust for the sampling schemes considered in their study given a limited degree of irregularity but that more extreme fluctuations in sampling period would cause misleading results. However, they elected to use a very narrow parameter range for the computation of

[a)]Electronic mail: michael.mccullough@research.uwa.edu.au

ordinal patterns with respect to the pattern length and focussed primarily on qualitative discussion of the frequency content of the signals and the Nyquist frequency obtained from Fourier transforms of the time series to explain their results.

Our investigation is intended to serve as an extension to the paper by Kulp *et al*. by virtue of several key distinctions between this and the former study. First, we study the benefits and potential limitations of exploring a larger parameter space for the ordinal pattern length. Second, and critically, we present the interpretation of our results and formulate our conclusions through the lens of nonlinear dynamical systems theory, which is not only novel and complementary to the existing study but also a necessary perspective in light of the much broader motivation: to detect quantitative evidence of deterministic dynamics in complex systems based on measured time series data. Differences in our approach that are of a technical nature will be detailed throughout this paper within the context in which they arise.

The remainder of this paper is structured as follows: Section II provides a brief introduction to the concept of ordinal patterns and forbidden patterns as a means for detecting determinism. Section III comprises the description of methodology, results, and discussions for three separate numerical investigations into the effect of sampling on the count of forbidden patterns as performed on time series from the Lorenz system and the hyper-chaotic Rössler system, namely: under-sampling (Sec. III A), random depletion of data (Sec. III B), and timing jitter (Sec. III C). Concluding remarks are given in Section IV.

II. COUNTING FORBIDDEN PATTERNS

Consider an arbitrary time series comprising a total of N temporally ordered observations from a continuous system. We begin by enumerating the set of ordinal patterns occurring in the time series using the Bandt and Pompe method.[10] This is done by taking a sliding window of length m along the time series and computing the rank order of data points within each window based on relative amplitude. The rank ordering of a given window is its ordinal pattern. For brevity, we omit further details of this mapping procedure and instead refer readers to the recent review paper by Amigó *et al*.[9] for a comprehensive definition of the process.

For a given window length, there are a total of $m!$ unique ordinal patterns that can possibly occur in our time series (see Figure 1, for example). It is intuitive that all $m!$ patterns will almost certainly occur in a time series generated by a stochastic process for $N \to \infty$. Perhaps less intuitive but pivotal in the context of this study is that for time series generated by deterministic dynamics there will exist a set of forbidden patterns which can never occur.[9] Using analytical arguments based on the concept of topological permutation entropy, it has been demonstrated that forbidden patterns will always exist for deterministic one dimensional piecewise monotone interval maps,[11] and numerical investigations suggest that forbidden patterns are also characteristic of continuous chaotic dynamics.[8] By virtue of this phenomenon, it is possible to use the count of the forbidden patterns to detect determinism in time series data. Complications arise from the fact that real data are finite; hence, an ordinal pattern that is admissible in the dynamics may not occur during the period of observation.[2] However, methods have been developed with this issue in mind, which are effective for the detection of determinism in relatively short, highly noisy data.[11,13]

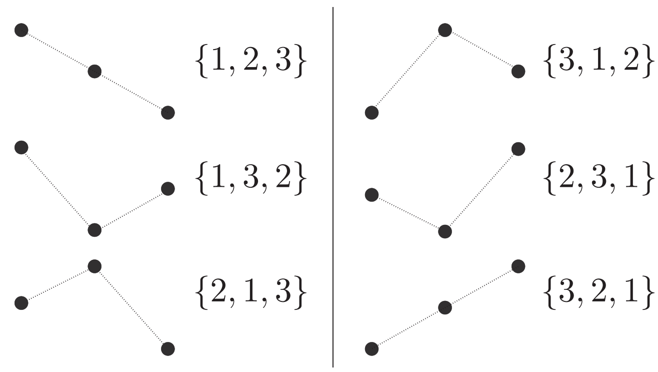


FIG. 1. The set of all possible ordinal patterns of length $m = 3$. Data points are ranked in descending order.

Having outlined the means by which the set of ordinal patterns can be computed from a time series and the theory underpinning the application of forbidden patterns to the detection of determinism in noisy time series, we return to the central question of this study: how reliable is the count of forbidden patterns from time series when the data are undersampled or irregularly sampled? For the remainder of this study, we will refer to the measure

$$\mathcal{P}_f^{(m)} = \frac{m! - |\mathbf{s}|}{m!}, \tag{1}$$

which is the relative number of forbidden patterns with respect to the total number of possible patterns dependent on m, where $\boldsymbol{s}$ is the set of ordinal patterns enumerated from the time series and $|.|$ denotes cardinality. This measure is an estimator of the relative number of true forbidden patterns, or alternatively, $1 - \mathcal{P}_f^{(m)}$ is the proportion of patterns admissible by the dynamics which would almost certainly occur as $N \to \infty$. In the following investigation, we test the robustness of $\mathcal{P}_f^{(m)}$ with respect to various sampling schemes for a range of m.

III. NUMERICAL INVESTIGATION

Before proceeding with the main line of inquiry, we present Figure 2 as a result which motivates the study of $\mathcal{P}_f^{(m)}$ for larger m. Kulp *et al*.[8] limited the scope of their investigations to $m \leq 5$. However, it is clear from Figure 2 that $\mathcal{P}_f^{(m)}$ appears to be a more robust measure against both noise and false forbidden patterns as m increases (provided $N \gg m!$). This is due to the outgrowth property which causes forbidden patterns to dominate (super-exponential growth) with respect to admissible patterns (exponential growth) as m increases for deterministic time series.[2] Moreover, while smaller m may be sufficient for detecting determinism in periodic dynamics and single scroll phase coherent chaos generated by the Rössler system, more complex chaotic time series from the two scroll Lorenz and Chua circuit attractors appear

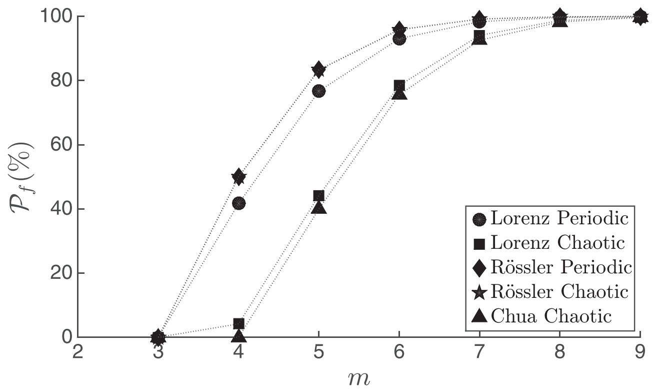
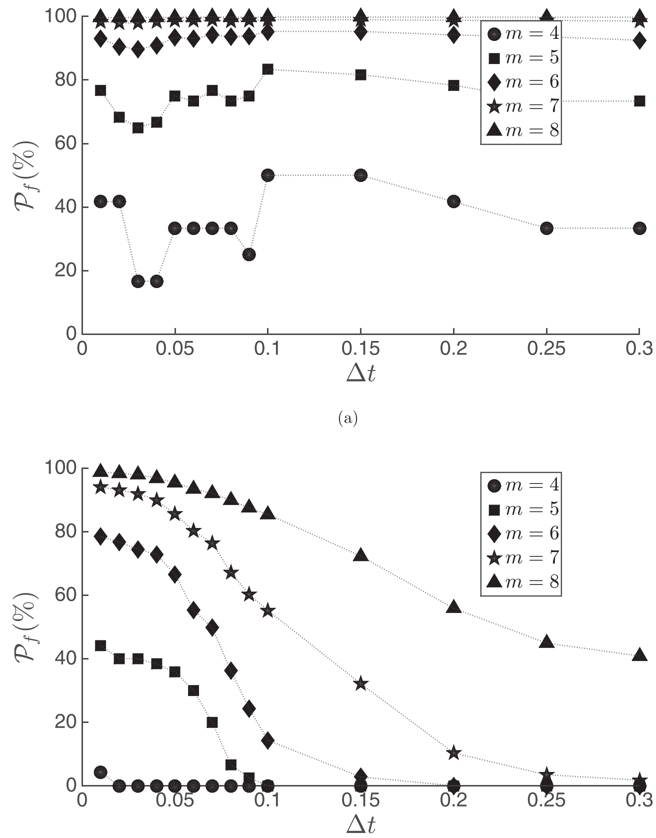


FIG. 2. Percentage of forbidden patterns against pattern order m in long ($N \gg m!$) highly sampled time series from periodic and chaotic regimes of archetypal chaotic systems.



(a)

(b)

FIG. 3. Percentage of forbidden patterns against sampling period Δt for regularly sampled time series generated by the Lorenz system in the (a) periodic and (b) chaotic regimes.

to require $m \geq 5$ before the measure would become a reliable estimator in the presence of even small amounts of noise.

The method of investigation used for this study generally follows the approach taken by Kulp *et al.*[8] but with several fundamental exceptions. First, we have elected to report the relative number of forbidden patterns as a percentage rather than an absolute count, thereby enabling easier comparison between results for different m. Second, Kulp *et al.* specified a fixed time series length N while varying the sampling period Δt. This meant that the total integration time length of each time series was directly proportional to the experimental variable Δt and changed with each simulation. Instead, we opt to fix the integration time length and allow N to vary. The practical analogy for our approach, distinct from the former, is that we are constrained to observe a system for some fixed time; hence, the system generates some fixed total amount of information, and we concern ourselves with the question of how highly or regularly the system be sampled to extract information of significant quantity and quality for $\mathcal{P}_f^{(m)}$ to be a reliable estimator.

We perform our investigation on highly sampled time series generated by the Lorenz system in both periodic and chaotic regimes, in line with Kulp *et al.*[8]—readers should refer to this paper for details of equations and system parameters. Time series were generated by solving the systems using a fourth-fifth order Runge-Kutta algorithm and transients were removed.

A. Under-sampling

To test the effect of under-sampling for a regular sampling period, we generate time series for total integration time length of 20 000 time units (approximately 6×10^5 cycles) for $0.01 \leq \Delta t \leq 0.3$ and then compute $\mathcal{P}_f^{(m)}$ for $3 \leq m \leq 8$. In the periodic case (Figure 3(a)), it can be observed that $\mathcal{P}_f^{(m)} \gg 0$ as expected and is stable for $6 \leq m \leq 8$. However, the value of the measure appears to be sensitive to Δt in a highly non-trivial manner when $m \leq 5$. This sensitivity is very likely due to aliasing effects as postulated by Kulp *et al.*, who also reported that when Δt is such that the sampling frequency $1/\Delta t$ is less than the Nyquist frequency, results should be discarded because the data are no longer sufficient to describe the waveform of the underlying dynamics.[8] On the latter point, we assert that that $\mathcal{P}_f^{(m)}$ will be non-zero for all Δt given regular sampling period and for sufficient m. This is because aliasing will result in a time series with a period greater than the original signal but one which is still strictly periodic; hence, forbidden patterns will still exist and be a characteristic indicator for the detection of determinism even as Δt becomes large.

For chaotic time series, results in Figure 3(b) show $\mathcal{P}_f^{(m)} \gg 0$ converging to zero for $m = 3, ..., 7$. This is best explained by the fact that time series points in each window become de-correlated with increasing Δt due to sensitivity to initial conditions imposed by the chaotic dynamics, and subsequently the time series will resemble a random process when Δt is sufficiently large. We can find an estimate for this value of Δt by computing the autocorrelation function and Lyapunov time of the time series, shown in Figure 4. Therefore, it should be expected that $\mathcal{P}_f^{(m)} > 0$ for $\Delta t \lesssim 0.28$ and $\mathcal{P}_f^{(m)} \approx 0$ for any larger Δt. Referring back to Figure 3(b), this suggests that selecting $m = 7$ provides the most robust $\mathcal{P}_f^{(m)}$ for the range of Δt where the time series retains evidence of the underlying determinism.

However, this also implies an anomaly in the results for $m = 8$ because $\mathcal{P}_f^{(m)} \gg 0$ and does not appear to be converging anywhere close to zero. We have included this curve to highlight the issue of false forbidden patterns and insufficient data. It is generally accepted that ordinal pattern statistics will be reliable if the length of data available meets the

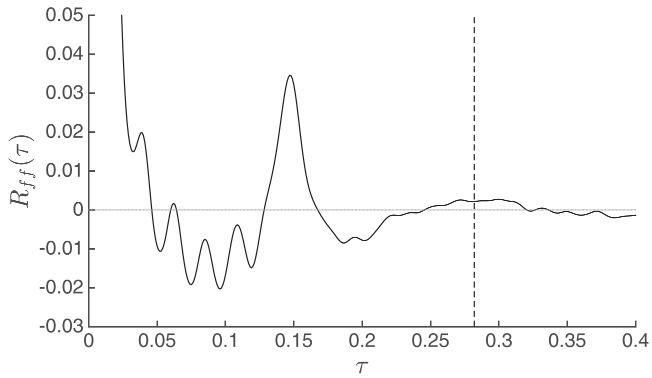


FIG. 4. The autocorrelation function and Lyapunov time (vertical dashed line) for the regularly sampled chaotic Lorenz time series. The Lyapunov time of the system which was estimated from the time series using the TISEAN[14] software package function *lyap_k*.

heuristic condition $N \gg m! + m - 1$ to ensure that it will be likely that most of the admissible patterns occur in the data.[2] In the case of the results for $m = 8$ shown in Figure 3(b), we have $N \approx 70\,000$ for $\Delta t = 0.28$ and $m! = 8! = 40\,320$, so this condition is not met as Δt approaches the Lyapunov time. Therefore, it is probable that the estimate of $\mathcal{P}_f^{(m)}$ for $m = 8$ includes many false forbidden patterns and, as such, is inaccurate. To confirm this, we compute $\mathcal{P}_f^{(m)}$ with $m = 8$ against $N \leq 10^6$ for several Δt either side of the Lyapunov time, as shown in Figure 5. These results show that $\mathcal{P}_f^{(m)}$ converges to zero (or very close to zero) with increasing N for $\Delta t > 0.28$ and is non-zero for Δt in the lower range where the time series exhibits a reasonable degree of correlation due to determinism. For the remainder of this study, we use $\Delta t = 0.05$ and 0.1 as our fundamental sampling periods (before depletion or jitter are applied to the data) because $\mathcal{P}_f^{(m)}$ is non-zero and stable up to $m = 8$ for $N > 10^5$.

Our key findings in this subsection are as follows. First, detection of determinism using only $\mathcal{P}_f^{(m)}$ will likely fail for chaotic time series when the sampling period exceeds the Lyapunov time or the point at which the autocorrelation function approaches zero. We note here, however, that this finding does not exclude the possibility that other more rigorous quantitative methods for detecting determinism using forbidden patterns[11,13] may still be effective for sparse irregularly sample time series. Second, selecting larger m will provide better detection of determinism when using forbidden patterns methods until the point where there are insufficient data to ensure accurate sampling of admissible ordinal patterns.

In order to verify our findings with other example systems, we repeated our numerical experiments using x-component time series generated by the Rössler system in an 8-periodic regime $(a, b, c) = (0.3848, 2, 4)$ and the four-dimensional (4D) Rössler system in a hyper-chaotic regime $(a, b, c, d) = (0.25, 3, 0.5, 0.05)$. The range of Δt values for simulations with these data are selected to ensure an approximate equivalence of sampling intervals with respect to the *points-per-mean-cycle* between the different time series data used in this study. We do this because one of the central aspects of these investigations is the relationship between the sampling frequency and the time scale of dynamical evolution, as highlighted in the study by Kulp *et al.*[8] Figure 6 shows the sensitivity of the relative count of forbidden patterns to the sampling period for the hyper-chaotic 4D Rössler time series. The results are different from those for the Lorenz time series, in that $\mathcal{P}_f^{(m)}$ remains non-zero for large Δt where $m \geq 5$, and even in the case of $m = 4$ where the measures does reach zero, it trends upwards again to a non-zero value. We postulate that the difference between the results for the two systems arises from their respective autocorrelation functions. As shown in Figure 4, the Lorenz time series decorrelates rapidly with values near zero. On the other hand, the autocorrelation function for the hyper-chaotic Rössler system is highly cyclic with slow decay. This is a reflection of the strongly quasi-periodic motion of the phase-space trajectory with respect to the one-dimensional projection onto the x-coordinate. Because ordinal patterns are robust to small changes in amplitude and the system has relatively consistent phase dynamics, the time series retains an element of its characteristic determinism to which ordinal patterns are sensitive, even for very sparse sampling, and hence why $\mathcal{P}_f^{(m)}$ is non-zero for large Δt. However, it is almost certain that $\mathcal{P}_f^{(m)}$ would eventually decrease. The dips in $\mathcal{P}_f^{(m)}$ for $m \leq 7$ are likely an aliasing phenomenon as also observed in the results for the periodic Lorenz time series (Figure 3(a)).

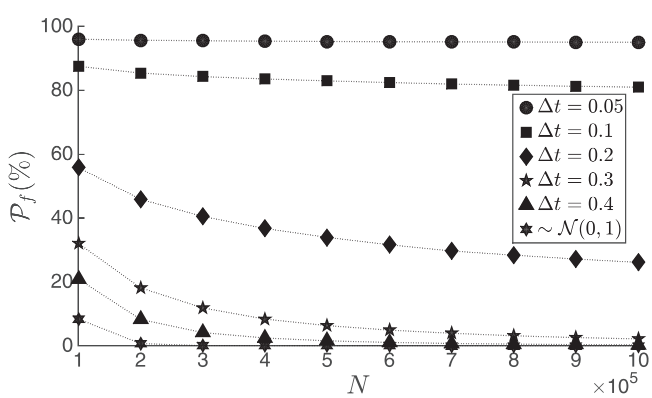


FIG. 5. The percentage of forbidden patterns against the length of the time series N for $m = 8$ (results for white noise are included for reference).

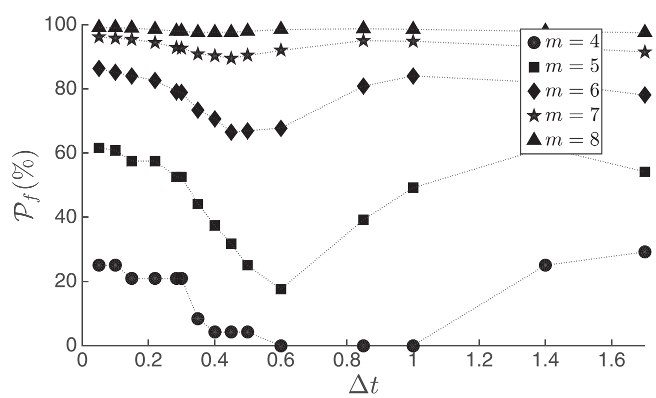


FIG. 6. Percentage of forbidden patterns against sampling period Δt for regularly sampled time series generated by the hyper-chaotic 4D Rössler system.

B. Random depletion

For this test, we generate time series for total integration time length of 20 000 time units for $\Delta t \leq 0.05$ and then randomly remove data points from the time series. Once more we have followed the methodology of Kulp *et al.*[8] but extend the computations to larger m.

Figure 7 shows $\mathcal{P}_f^{(m)}$ with respect to the percentage depleted ($d_{\%}$). Beginning at the leftmost point of the figure, one observes that $\mathcal{P}_f^{(m)}$ decreases with $d_{\%}$ for both periodic and chaotic dynamics. The count of forbidden patterns decreases because random depletion introduces a stochastic component in the dynamics by way of random distortion of the temporal correlation between elements in each m-length window of successive data points from which the ordinal patterns are computed. This eventually results in time series with no forbidden patterns, as can be observed in the curves for $m = 5, ..., 7$, at which point $\mathcal{P}_f^{(m)}$ is no longer a reliable criterion for determinism.

The first feature of these results which is new in this study is that all of the curves ultimately trend upward to $\mathcal{P}_f^{(m)} = 100\%$, most notably for $m = 8$ where the forbidden pattern count trends upward at $d_{\%} \approx 80\%$ before it is able to reach zero. This is due to false forbidden patterns becoming prevalent such that the condition for reliable sampling of ordinal patterns $N(1 - d_{\%}) \gg m!$ is no longer met. For example, for $d_{\%} = 90\%$ we have $0.1N = 40\,000 \approx 8!$ and therefore $d_{\%}$ must be far less than this level for $\mathcal{P}_f^{(m)}$ to be a usable measure when $m = 8$.

The second new finding in this study pertaining to random depletion is that, as per the under-sampling test, the results for both periodic and chaotic time series show that $\mathcal{P}_f^{(m)}$ is higher for larger m and therefore provides a more robust measure for the detection of determinism. This can be explained by examining the worst case for $\mathcal{P}_f^{(m)}$ affected by the process of random depletion. For simplicity, assume very long data and a level of depletion such that $N(1 - d_{\%}) \gg m!$. Then assume the case that for all m-length windows of the data, removing a single data point from within a window will result in a new ordinal pattern that was forbidden before the removal of the said data point and that the pattern originally corresponding to this window is also counted somewhere else in the data. Now consider $Nd_{\%}$ points depleted from the data each spaced by a minimum of $m - 1$ data points such that the maximal number of new patterns, equal to $Nmd_{\%}$, is observed. Therefore, forbidden patterns can only be observed if $m! - Nmd_{\%} > 0$. This condition is not intended to serve as a heuristic for electing to use forbidden patterns as a test for determinism because in that capacity

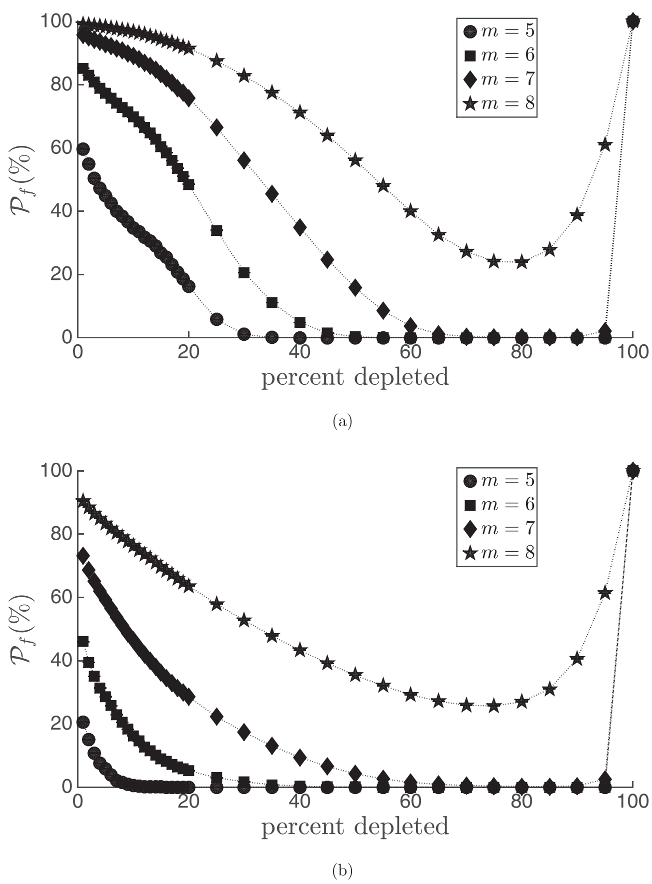


FIG. 7. Percentage of forbidden patterns against the percentage of the observations that have been randomly removed from the data for time series generated by the Lorenz system in the (a) periodic and (b) chaotic regimes. The data were sampled with regular period $\Delta t = 0.05$ prior to the depletion process. Each datum is the mean of 100 realisations of the simulation. Variance across all simulations is less than 2.4%.

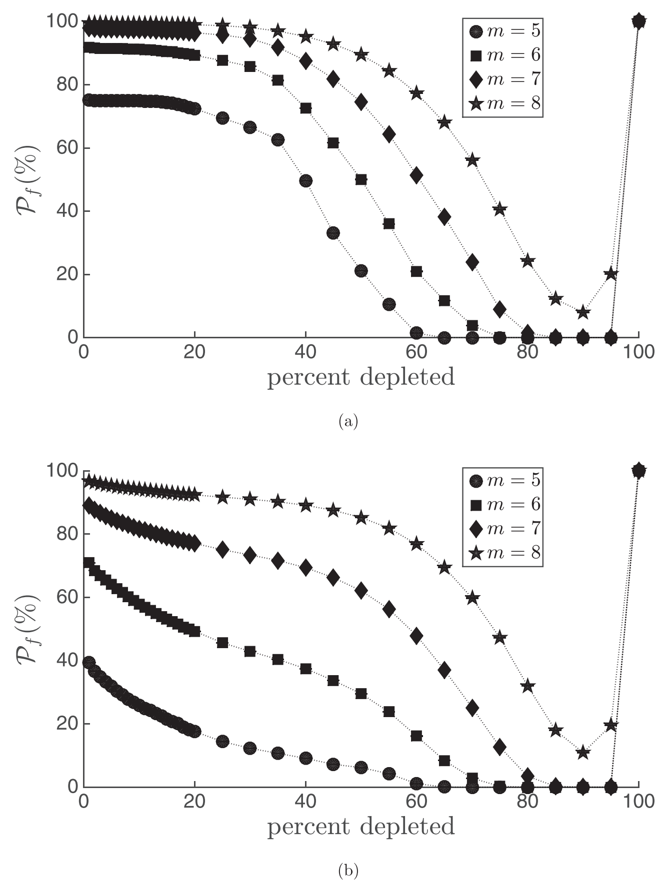


FIG. 8. Percentage of forbidden patterns against depletion percentage for time series generated by the (a) 8-periodic Rössler and (b) hyper-chaotic 4D Rössler time series. The data were sampled at regular periods (a) $\Delta t = 0.31$ and (b) $\Delta t = 0.285$, respectively, prior to the depletion process. Each datum is the mean of 100 realisations of the simulation. Variance across all simulations is less than 0.2%.

it is excessively conservative. However, it implies that for a given length of data and percentage of depletion, using larger m should provide a more reliable count of forbidden patterns where sufficient data are available.

Figure 8 shows the corresponding results for the 8-periodic Rössler and the hyper-chaotic 4D Rössler experiments. The Δt value used is such that we sample 20 points per mean cycle in both cases. Evidently, the results are consistent with our previous analysis for the Lorenz data. We have elected to include the plot for the 8-periodic regime to highlight two observations. First, the large $\mathcal{P}_f^{(m)}$ count which characterises periodic systems. Second, the robustness of the count even at significant depletion percentages, as the very low rate of decrease for $d_{\%} \leq 40\%$ in Figure 8(a) illustrates.

C. Timing jitter

In this final test, we investigate $\mathcal{P}_f^{(m)}$ for time series with simulated timing jitter applied to the sampling period. This is achieved by defining a vector of sampling times with regular period and perturbing each sample time by realisations of a random variable with a normal distribution of zero mean. The intensity of the jitter is controlled by the standard deviation σ. If a sampling time index is perturbed to the extent that it overlaps with adjacent indices, then the vector is sorted to recover temporal ordering. The Runge-Kutta solver returns observations for the periodic and chaotic Lorenz systems at the times specified by the sampling vector for total integration time length of 20 000 time units. Note that $N \gg m!$ holds for these tests.

The results for the periodic case are shown in Figure 9. It can be observed that $\mathcal{P}_f^{(m)}$ appears to converge to a steady value when $\sigma \approx \Delta t$. The increasing jitter perturbs the temporal correlation between adjacent data points resulting in more admissible patterns until $\sigma > \Delta t$ when the sampling scheme is essentially random with constant average density on the time scale Δt. This is likely to be an unrealistic level of jitter for most applications; nonetheless, in this scenario the primary factor affecting the count of forbidden patterns is the average density of sampling. As can be seen by comparing Figure 9(a) with Figure 9(b), larger sampling periods make $\mathcal{P}_f^{(m)}$ less reliable in the presence of the timing jitter. However, the key finding is that $\mathcal{P}_f^{(m)}$ maintains non-zero values for σ as large as $2\Delta t$ and beyond when m is sufficiently large. The prior work of Kulp *et al.*[8] only provides results for the jitter timing test for $m = 5$. For this choice of m, $\mathcal{P}_f^{(m)}$ is only non-zero in the periodic time series for jitter with $\sigma < 0.5$ when $\Delta t = 0.05$.

Results are similar in the chaotic case as observable in Figure 10. The estimator $\mathcal{P}_f^{(m)}$ remains a reliable criterion for determinism for sufficiently large m. On the other hand, for larger Δt the count of forbidden patterns is severely diminished in the presence even for small σ (Figure 10(b)) despite this sampling period providing reliable results for regular sampling. This is because the correlation between adjacent data points in a chaotic time series disintegrates rapidly as Δt increases due to exponential divergence of trajectories as discussed in Section III A. Under these dynamical conditions, even small perturbations to temporal correlation can be

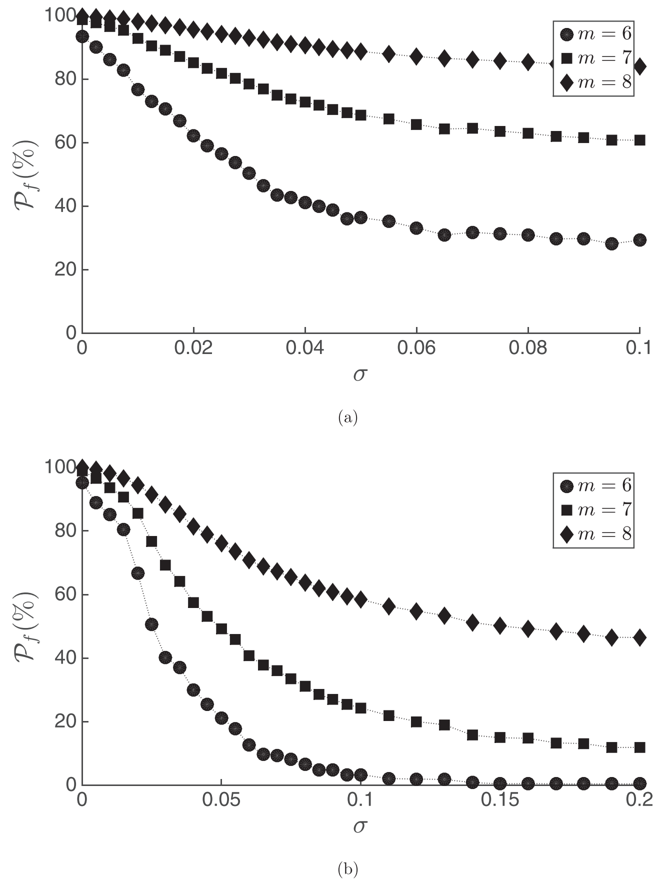


FIG. 9. Percentage of forbidden patterns against the timing jitter for time series generated by a periodic Lorenz system. In this experiment, the underlying regular sampling period was set to (a) $\Delta t = 0.05$ and (b) $\Delta t = 0.1$, and the degree to which the regular sampling period was perturbed was governed by $\sim\mathcal{N}(0, \sigma^2)$.

significant. Finally, as was clear for periodic time series, it is reasonable to postulate from these curves that the density of sampling, rather than the regularity of the sampling period, should be the primary concern when assessing the likelihood of obtaining a reliable $\mathcal{P}_f^{(m)}$ from experimental data affected by timing jitter.

Figure 11 depicts the results for the hyper-chaotic 4D Rössler experiment. All key findings are consistent with those for the Lorenz time series. We observe a saturation to a limiting value once $\sigma \simeq \Delta t = 0.285$ for $m \geq 7$ and non-zero values for all $\sigma = 2\Delta t$ and beyond regardless of m. The estimator $\mathcal{P}_f^{(m)}$ retains its reliability in detecting determinism and the regularity of sampling period seems to be of secondary concern in comparison to the density of sampling. Note that the sampling frequency of the underlying (unperturbed) regular grid corresponds to 20 points per mean cycle, which is sufficiently high as shown in Figure 6.

IV. CONCLUDING REMARKS

In this study, we have investigated the robustness of the count of forbidden patterns as a criterion for the detection of determinism with respect to data that have been undersampled or irregularly sampled by a process of random depletion and timing jitter. For each flawed sampling

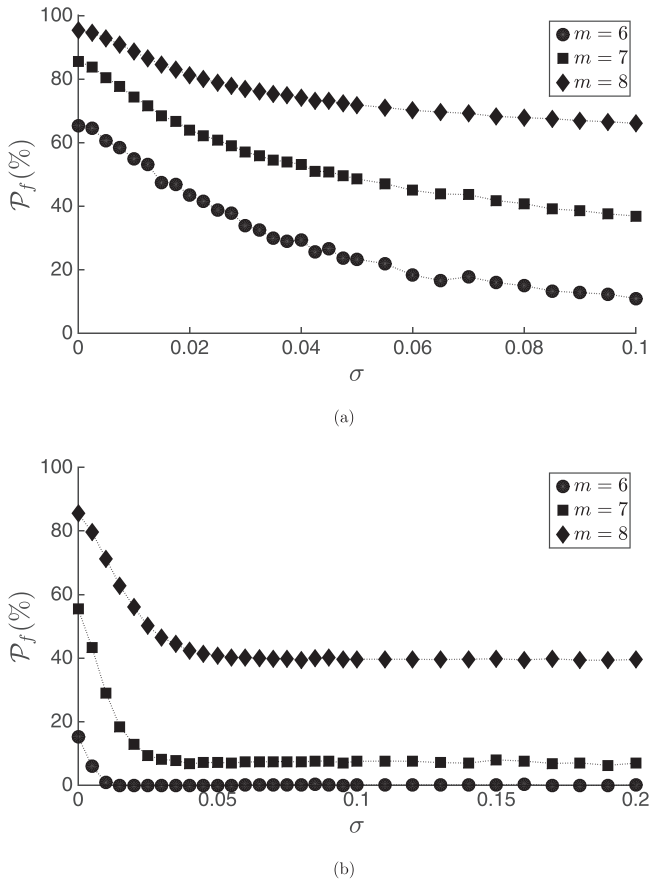


FIG. 10. Percentage of forbidden patterns against the timing jitter for time series generated by a chaotic Lorenz system. In this experiment, the underlying regular sampling interval was set to (a) $\Delta t = 0.05$ and (b) $\Delta t = 0.1$, and the degree to which the regular sampling period was perturbed was governed by $\sim\mathcal{N}(0, \sigma^2)$.

scheme, we generated continuous model time series from data from the Lorenz system in both periodic and chaotic regimes and the Rössler system in 8-periodic and hyper-chaotic regimes. From these data, we computed the relative number of forbidden patterns for ordinal patterns up to $m = 8$. Our results demonstrated that the count of forbidden patterns is robust under certain conditions for all of the sampling schemes that we tested.

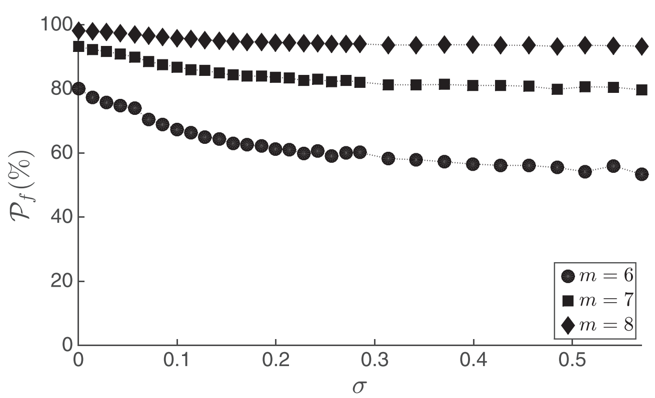


FIG. 11. Percentage of forbidden patterns against the timing jitter for time series generated by the hyper-chaotic 4D Rössler system. In this experiment, the underlying regular sampling period was set to $\Delta t = 0.285$, and the degree to which the regular sampling period was perturbed was governed by $\sim\mathcal{N}(0, \sigma^2)$.

In the case of under-sampling for periodic time series, we found that the count of forbidden patterns remains non-zero for significantly under-sampled data and showed that it will remain non-zero for any regular sampling interval given sufficient m. For chaotic time series, successive time series points will rapidly de-correlate as the sampling interval increases due to the sensitivity to initial conditions that is inherent in chaotic systems. When the sampling interval is too large, the time series resembles a random process and, hence, forbidden patterns will no longer occur. We found it necessary for the sampling interval to be in the range where correlation can still clearly be observed from the autocorrelation function of the time series. Where data have been randomly depleted from the time series we observe that the count of forbidden patterns will become unreliable when the length of data available approaches the number of possible permutations, as governed by the window length parameter m. This finding applies to all of our results—unsurprisingly because this is a well known condition for sampling ordinal patterns. For time series affected by timing jitter, our results indicate that it is the average density of sampling rather than the degree of jitter, which determines the reliability of the count of forbidden patterns.

The central, overarching finding in this investigation is, however, that the window length m should be chosen as large as possible for a given length N of data available. Broadly speaking, our results agree with those of Kulp *et al.*[8] yet we found that the count of forbidden patterns was significantly more robust when m was chosen larger in all instances. Furthermore, we provided a simple analytical argument for the importance of selecting m as large as possible for the case of random depletion.

There are, of course, numerous other models for irregular sampling schemes, including Poisson and power law distributions, which may be closer to reality in certain applications such as the analysis of pre-historic climate data. We refer readers to the second paper in this series where we continue our investigation of the reliability of forbidden patterns in data that have been irregularly sampled based on these and other distributions.[15]

ACKNOWLEDGMENTS

We would like to extend our thanks to Dr. Christopher Kulp and his co-authors for their work, as cited, which served as the primary motivation for this research. M.S. is supported by the Australian Research Council Discovery Project No. DP 140100203.

[1]E. Bradley and H. Kantz, Chaos **25**, 097610 (2015).
[2]J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, Europhys. Lett. **79**, 50001 (2007).
[3]J. D. Scargle, Astrophys. J. **263**, 835 (1982).
[4]S. Suzuki, Y. Hirata, and K. Aihara, Int. J. Bifurcation Chaos **20**, 3699 (2010).
[5]R. Xia, O. Odemuyiwa, F. Murgatroyd, J. Gill, M. Malik, and A. J. Camm, in *Proceedings of Computers in Cardiology* (1992), pp. 307–310.

[6]G. G. Berntson and J. R. Stowell, Psychophysiology **35**, 127 (1998).
[7]T. M. Souders, D. R. Flach, C. Hagwood, and G. L. Yang, "The effects of timing jitter in sampling systems," IEEE Trans. Instrum. Measurement **39**(1), 80–85 (1990).
[8]C. W. Kulp, J. M. Chobot, B. J. Niskala, and C. J. Needhammer, Chaos **26**, 023107 (2016).
[9]J. M. Amigó, K. Keller, and V. A. Unakafova, Philos. Trans. R. Soc., A **373** (2014).
[10]C. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002).
[11]J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, Int. J. Bifurcation Chaos **20**, 2915 (2010).
[12]O. A. Rosso, L. C. Carpi, P. M. Saco, M. G. Ravetti, A. Plastino, and H. A. Larrondo, Physica A **391**, 42 (2012).
[13]J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, Europhys. Lett. **83**, 60005 (2008).
[14]R. Hegger, H. Kantz, and T. Schreiber, Chaos **9**, 413 (1999).
[15]K. Sakellariou, M. McCullough, T. Stemler, and M. Small, Chaos **26**, 123104 (2016).